CLAIMS

1.   **(Currently Amended)** A processor-readable medium comprising processor-executable instructions for personalizing karaoke, the processor-executable instructions comprising instructions for performing a method, the method comprising:

~~~~~~~~obtaining music;

~~~~~~~~obtaining lyrics corresponding to the music from a file;

~~~~~~~~selecting a visual content according to the content, a user's preference, and a type of music with which the visual content is to be aligned;

~~~~~~~~segmenting music to produce a plurality of music sub-clips, ~~wherein the segmenting establishes boundaries between the music sub-clips at beat positions within the music, the beat positions being located according to a rhythm or a tempo of the music, or at onset positions within the music when beat positions are not obvious during a portion of the music, the onset positions being initiations of distinguishable tones of the portion of the music, wherein lengths of the sub-clips are shorter than a maximum of sub-clips length~~;

segmenting a visual content to produce a plurality of sub-shots at a maximum peak of a frame difference curve, wherein the visual content presents a story line and the segmenting is repeated until lengths of all sub-shots are shorter than a maximum of sub-shot length, the maximum of sub-short length being a little longer in duration than the maximum of music sub-clips to facilitate the sub-short being truncated to equal a length of an aligned music sub-clip in a next step;

selecting sub-shots from the plurality of sub-shots, the selecting comprising:

filtering sub-shots from within the plurality of sub-shots according to importance and quality, the filtering sub-short from within the plurality of sub-shorts according to importance comprising:

calculating an attention/importance index of each frame of the sub-shot based on a plurality of factors including object motion, camera motion, specific objects, and audio, if any, associated with the frame;

calculating an attention/importance index of the sub-short by averaging the attention/importance index of each frame of the sub-short; and

selecting the sub-shots by comparing the attention index of each sub-shot; and

selecting sub-shots such that they are uniformly distributed along a time line of ~~within~~ the visual content to preserve the story line of the visual content;

aligning sub-shots with music sub-clips, the aligning comprising:

automatically shortening one or more of the plurality of sub-shots to a length of a corresponding music sub-clip from within the plurality of music sub-clips; and

resolving differences in the number of sub-shorts and the number of music sub-clips;

~~obtaining lyrics corresponding to the music from a file;~~

coordinating delivery of the lyrics with the music using timing information contained within the file; and

displaying at least some of the plurality of sub-shots as a background to lyrics associated with the plurality of music sub-clips, the displaying comprising:

merging the selected sub-shots into scenes by a plurality of grouping methods, the methods including:

merging the sub-shorts by similarity; and

merging based on a time-code or timestamp of the sub-shots; and

producing a number of effects at transitions of the plurality of sub-shots.

2 - 4. **(Cancelled)**

5. **(Previously Presented)** The processor-readable medium as recited in claim 1, wherein filtering the plurality of sub-shots according to quality comprises:

examining color entropy within each of the plurality of sub-shots for indications of diffusion of color; and

if color entropy is low, analyzing each of the plurality of sub-shots to detect motion more than a threshold indicating interest and less than a threshold indicating low camera and object movement; and

selecting sub-shots having acceptable motion and color entropy scores.

6.   **(Previously Presented)** The processor-readable medium as recited in claim 1, wherein filtering the plurality of sub-shots according to importance comprises:

evaluating frames within a sub-shot according to attention indices; and

averaging the attention indices for the frames to determine if the sub-shot should be included or excluded.

7.   **(Cancelled)**

8.   **(Previously Presented)** The processor-readable medium as recited in claim 1, wherein each sub-shot comprises a segment of video of at least a predetermined length based on a length of the music sub-clips and segmented based on a magnitude of difference between adjacent frames .

9.   **(Cancelled)**

10.   **(Previously Presented)** The processor-readable medium as recited in claim 1, wherein selecting sub-shots comprises:

evaluating color entropy, camera motion, object motion and object detection; and

selecting the important sub-shots based on the evaluation.

11.   (Previously Presented) The processor-readable medium as recited in claim 1, wherein selecting sub-shots comprises:

evaluating normalized entropy of the sub-shots along a time line of video from which the sub-shots were obtained.

12.   (Previously Presented) The processor-readable medium as recited in claim 1, wherein segmenting visual content comprises assigning photographs to be sub-shots.

13.   (Previously Presented) The processor-readable medium as recited in claim 12, wherein assigning photographs to be sub-shots comprises:

rejecting photographs having problems with quality; and

rejecting photographs within a group of very similar photographs wherein a photo within the group has been selected.

14.   (Currently Amended) The processor-readable medium as recited in claim 12, wherein assigning photographs to be sub-shots comprises:

converting at least one of the photographs to video, wherein camera angles change, zoon and pan the photograph.

15.   (Currently Amended) The processor-readable medium as recited in claim 1, wherein the visual content comprises one or more home videos and or photographs in

digital formats<u>, in an event that both video and photograph are used, each photograph is regarded as a video shot</u>.

16.   **(Canceled)**

17.   **(Previously Presented)** The processor-readable medium as recited in claim 1, wherein segmenting music into the plurality of music sub-clips comprises bounding music sub-clip length according to:

minimum length = min{max{2* tempo, 2}, 4} and

maximum length = minimum + 2.

18.   **(Previously Presented)** The processor-readable medium as recited in claim 1, wherein segmenting the music comprises:

establishing music sub-clips' length within a range of 3 to 5 seconds.

19.   **(Currently Amended)** The processor-readable medium as recited in claim 18, wherein segmenting the music comprises:

establishing boundaries for the music sub-clips at sentence breaks <u>in lyrics</u>.

20.   **(Cancelled)**

21.   **(Previously Presented)** A processor-readable medium as recited in claim 1, wherein obtaining the lyrics comprises sending the file over a network to a karaoke device as a part of a pay-for-play service.

22. **(Previously Presented)** The processor-readable medium as recited in claim 1, wherein the method further comprises:

querying a database of songs by humming a portion of a desired song; and

selecting the desired song from among a number of possibilities suggested by an interface to the database.

23. **(Currently Amended)** A processor-readable medium comprising processor-executable instructions for integrating lyrics, music and video content suitable for karaoke, the processor-executable instructions comprising instructions for performing a method, the method comprising:

receiving a request for a file associated with a specified song, wherein the file comprises: music, lyrics, and timing values associated with the lyrics;

fulfilling the request for the file by sending the file associated with the specified song;

segmenting the music to produce a plurality of music sub-clips, wherein the segmenting establishes boundaries between the music sub-clips at beat positions within the music, wherein the beat positions are located according to a rhythm or a tempo of the music;

segmenting a visual content representing a story line to produce a plurality of sub-shots of a length corresponding music sub-clips from the plurality of music sub-clips, such that the plurality of sub-shots are uniformly distributed along a time line of within the visual content to preserve the story line of the visual content; and

outputting the plurality of music sub-clips together with corresponding sub-shots of visual content, wherein the visual content is configured as a background to the lyrics associated with the music sub-clips.

24. **(Previously Presented)** A processor-readable medium as recited in claim 23, wherein obtaining the lyrics comprises sending the file over a network to a karaoke device.

25. **(Currently Amended)** A personalized karaoke device, comprising:

a music analyzer configured to segment a music to produce a plurality of music sub-clips, wherein the segmenting establishes boundaries between the music sub-clips at beat positions within the music of a song, wherein the beat positions are located according to a rhythm or tempo of the music;

a visual content analyzer configured to define and select visual content sub-shots, wherein the visual content analyzer is configured to select sub-shots of greater importance consistent with creating a uniform distribution of the sub-shots over a runtime of a source video, wherein the source video presents a story line and the sub-shots preserve the story line of the source video;

a lyric formatter configured to time delivery of syllables of lyrics of the song; and

a composer configured to:

assemble the music-sub clips with the visual content sub-shots;

adjust length of the sub-shots to correspond to the music sub-clips; and

superimpose the syllables of the lyrics of the song over the sub-shots.

26. **(Original)** The personalized karaoke device of claim 25, wherein the music analyzer is configured to segment the song with a strong onset between each of the music sub-clips.

27. **(Original)** The personalized karaoke device of claim 25, wherein the music analyzer is configured to segment the song with a beat between each of the music sub-clips.

28. **(Original)** The personalized karaoke device of claim 25, wherein the music analyzer is configured to segment the song automatically into sub-clips, each having a duration that is a function of song tempo.

29. **(Original)** The personalized karaoke device of claim 25, wherein the visual content analyzer is configured to segment video into sub-shots.

30. **(Original)** The personalized karaoke device of claim 25, wherein the visual content analyzer is configured to access folders of home video and photographs containing content from which the sub-shots are derived.

31. **(Original)** The personalized karaoke device of claim 25, wherein the visual content analyzer is configured to assemble still photographs, each of which is a sub-shot.

32. **(Original)** The personalized karaoke device of claim 25, wherein the visual content analyzer is configured to select from among sub-shots according to ranked importance, wherein importance is gauged by detection of color entropy, detection of object motion within the sub-shot, detection of camera motion during the sub-shot, and/or detection of a face within the sub-shot.

33. **(Original)** The personalized karaoke device of claim 25, wherein the visual content analyzer is configured to filter out sub-shots having low image quality as measured by low entropy and low motion intensity.

34. **(Previously Presented)** The personalized karaoke device of claim 25, wherein the visual content analyzer is configured to define sub-shots from visual content comprising photographic and video content .

35. **(Previously Presented)** The personalized karaoke device of claim 34, wherein the visual content analyzer is configured to reject photographs of low quality by detecting over and under exposure, overly homogeneous images and blurred images.

36.   **(Original)** The personalized karaoke device of claim 25, wherein the visual content analyzer is configured to organize photographs by date of exposure and by scene, thereby obtaining photographs having a relationship.

37.   **(Previously Presented)** The personalized karaoke device of claim 36, wherein the visual content analyzer is configured to reject photographs which are members within a group of very similar photographs, wherein one of the group has already been selected.

38.   **(Original)** The personalized karaoke device of claim 25, wherein the visual content analyzer is configured to:

    detect an attention area within a photograph; and

    create a photo to video sub-shot based on the attention area, wherein the video includes panning and/or zooming.

39.   **(Original)** The personalized karaoke device of claim 25, wherein the lyric formatter is configured to consume a file detailing timing of each syllable and each sentence of the lyrics.

40.   **(Currently Amended)** An apparatus comprising:

    means for creating music sub-clips by segmenting the music to define boundaries between the music sub-clips at beat positions within a song, wherein the beat positions are located according to a rhythm or tempo of the music;

means for defining and selecting visual content sub-shots from a visual content, such that the sub-shots are uniformly distributed <u>along a time line of</u> ~~within~~ the visual content, wherein the visual content presents a story line and the sub-shots preserve the story line <u>of the visual content</u>;

means for timing delivery of syllables of lyrics of the song; and

means for assembling the music sub-clips with the visual content sub-shots, and to adjust length of the sub-shots to correspond to length of the music sub-clips, and to superimpose the syllables of the lyrics of the song over the sub-shots.

41.   **(Original)** The apparatus of claim 40, wherein the means for defining and selecting visual content sub-shots is a video analyzer configured to segment video into sub-shots.

42.   **(Original)** The apparatus of claim 40, wherein the means for defining and selecting visual content sub-shots is a video analyzer configured to access folders of home video and photographs containing content from which the sub-shots are derived.

43.   **(Original)** The apparatus of claim 40, wherein the means for defining and selecting visual content sub-shots is a video analyzer configured for:

detecting an attention area within a photograph; and

creating a photo to video sub-shot based on the attention area, wherein the video includes panning and zooming.

44. **(Original)** The apparatus of claim 40, wherein the means for timing delivery of syllables of lyrics of the song is a lyric formatter configured for consuming a file detailing timing of each syllable and each sentence of the lyrics and for rendering the lyrics syllable by syllable.

45. **(Previously Presented)** The apparatus of claim 40 further comprising:

means for displaying assembled visual content comprising sub-shots with music sub-clips; and

wherein:

the means for defining and selecting visual content sub-shots, such that the sub-shots are uniformly distributed within the visual content is further configured for selecting uniformly distributed sub-shots via evaluating normalized entropy of the sub-shots along a time line of visual content from which the sub-shots were obtained; and

the means for displaying the assembled visual content comprising sub-shots with music sub-clips is configured such that displaying the assembled visual content preserves a storyline as represented by the visual content.

46. **(New)** The processor-readable medium as recited in claim 1, wherein the segmenting establishes boundaries between the music sub-clips at beat positions within the music, the beat positions being located according to a rhythm or a tempo of the

music, or at onset positions within the music when beat positions are not obvious during a portion of the music, the onset positions being initiations of distinguishable tones of the portion of the music, wherein lengths of the sub-clips are shorter than a maximum of sub-clips length.

47.   **(New)**   The processor-readable medium as recited in claim 15, wherein the one or more photographs are grouped into three tiers including: a date that the photograph is taken, a scene within the photograph, and whether the photo is a member of a group of very similar photographs,

wherein the scene represents a group of photographs that, while not as similar as those which fall under the group of very similar photos, are taken at a same time and place.

48.   **(New)**   The processor-readable medium as recited in claim 47, wherein the date and scene are used to determine the number of effects at transition of the one or more photos and photos fall within a group of very similar photos are filtered out.

49.   **(New)**   The processor-readable medium as recited in claim 48, wherein the photographs are firstly grouped into a top-tier based on the date, then a hierarchical clustering algorithm with different similarity thresholds is used to group the lower two layers,

wherein the photographs with a lower degree of similarity are grouped together as the scene.

50. **(New)** The processor-readable medium as recited in claim 1, wherein the number of effects at transitions of the plurality of sub-shots are selected randomly in a plurality of specific effect sets or determined by a style.

51. **(New)** The processor-readable medium as recited in claim 50, the style includes a day-by-day style, wherein a title is added when a new day starts before a first sub-shot of the day to illustrate the creating of the sub-shots coming next.

52. **(New)** The processor-readable medium as recited in claim 50, the style includes an old movie style, wherein sepia tone or grayscale effect is added on the sub-shots.